Least Squares Problems

Math 647

1. Suppose that a curve of the form $y = ax + \frac{b}{x}$ is fit to data points $(x_1, y_1), \cdots, (x_n, y_n)$ using a least squares fit. Find formulas for a and b in terms of the data points.

2. The average high temperatures in a given month in College Station is given in the following table:

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
61	66	73	79	85	92	96	96	91	82	71	63

   (a) Make a scatter plot of this data. (Take January to be 1, February to be 2, and so on.)

   (b) Using least squares, fit a curve of the form $a + b\sin\left(\frac{\pi}{6}t + c\right)$ to the data. (The factor of $\frac{\pi}{6}$ is so that the curve will be periodic of period 12.) Use your scatter plot in part a to give reasonable intervals for the variables in fsolve. Plot your least squares curve on top of the scatter plot.

3. The following table is taken from a study on the growth of duckweed in a pond. The top row is the day, the bottom row is the number of fronds of duckweed present.

0	1	2	3	4	5	6	7	8	9	10	11	12	13
100	127	171	233	323	452	654	918	1406	2150	2800	4140	5760	8250

   (a) Make a scatter plot of this data.

   (b) Make a scatter plot of the natural log of the number of fronds vs. days (to take the ln of each element in an array, say FROND, use the command "map(ln,FROND)". The result will be an array of the logs of the entries.)

   (c) Find the least squares line to fit the log of the number of fronds vrs days. What exponential is this predicting for the number of fronds as a function of the day?

   (d) Now do least squares fit of an exponential directly to the data, as I did in the lecture notes for the population of the US. Use your answer to part c to help you put reasonable intervals into fsolve for the parameters.

(e) Plot the exponential from part c and the exponential from part d on top of your scatter plot. (Use different colors for the different curves, so that you can tell which is which.)

4. It's my suspicion that you can make a pretty good prediction of how a student will do in a Calc I based on their performance on the first two homework assignments. For 10 students from a class of mine a while back, I have the following data (the homework was out of 20 points).

HW#1	HW#2	Final percent
13.5	17.75	80.6
13	8	66.3
14.5	15.25	54.3
13	14.5	76.5
18.5	17.25	86.0
19.5	14.5	77.6
16.5	12.75	84.1
12	15.25	81.4
18.5	15.75	81.9
16	15.75	91.2

Use a linear least squares regression to approximate the final grade in terms of the first two homework grades. In other words, if f is the final grade, x is the first homework score, and y is the second homework score, find a, b, and c so that $a + bx + cy$ is the best fit to f in the sense of least squares. Use this to predict the final grade for a student who gets a 17 and a 16.75 on the first two assignments.